知识工程中现代语义分析方法的应用及检讨

「知識工程」（人類の知恵集積プロジェクト）における現代の語義の分析方法——その応用と検討

霍四通

本文分别结合知网（HowNet）和WordNet讨论知识工程中最常用的两种语义分析方法，即义素分析和语义场分析的方法。知网和WordNet分别是目前应用最广泛的汉语和英语的知识库。知网是以汉语和英语的词语所代表的概念为描述对象，以揭示概念与概念之间以及概念所具有的属性之间的关系为基本内容的常识知识库，1999年上网发布，并不断更新，最新版为2008年版。本文以1.0版为讨论对象。WordNet是由美国George A. Miller主持完成的一个大型的英语知识库（a large lexical database of English），迄今也不断更新，本文也主要以较早（1997年）发布的1.6版为考察对象。

一、知网的建设和义素分析

义素是构成词义的最小意义单位，也就是词义的区别特征。如“单身汉”一词，一般认为可将它分解成“没有结婚的”、“成年的”、“男人”三个义素。这三个义素分别代表三个属性，“单身汉”的外延就通过内涵为这三个原子概念的类别的交集确定（Lyons1977:319）。

在知网中，所有的词语定义都是根据“义原”得出的。知网对义原作了这样的阐述：“大体上说，义原是最基本的，不易于再分割的意义的最小单位。例如：‘人’虽然是一个非常复杂的概念，它可以是多种属性的集合体，但我们也可以把它看作为一个义原。”（董振东，董强1999）知网就是基于这样的一个信念来建设知识系统的：所有的概念都可以分解成各种各样的义原。同时，这些义原又是有限的，是一个有限的集合，但根据其中的义原能组合成一个无限的概念集合。因此把握了这一有限的义原集合，就可以利用它来描述概念之间的关系以及属性与属性之间的关系，就有可能建设出一个知识系统。因此说，知网的义原就是“义素分析”中的“义素”。

对照义素分析的一般方法，我们可以更好的理解知网的建设方法。

首先是义素的来源。义素总是靠内省的方法获得的。这种靠语感来分析义素的方法一般来说也是可靠的，因为语感并不是个人的、纯主观的东西，它是说话人在长期使用某一语言的过程中，逐渐形成的一种能力；只要研究人或被咨询人对被分析的对象十分熟悉，就可以试着采用这种比较方法（贾彦德1999：64）。但更方便的做法是依靠词典的定义。一般的词典在揭示词义时往往把它的属性排列出来，首先通过类属关系同别的非本类事物相区别，再根据某些特征同本类内部其他事物相区别。例如：

灌木——矮小而丛生的木本植物。

乔木——树干高大，主干和分支有明显区别的木本植物。

定义中的“丛生”就是主干和分支无明显区别。如果用“+”表示有此特征，“－”表示无此特征，可以把这两个词表示如下：

灌木——＋[矮小]＋[丛生]＋[木本]＋[植物]

乔木——－[矮小]－[丛生]＋[木本]＋[植物]

方括号内的特征正是构成这些词义的最小单位，也就是它们的义素。

其次是语义的颗粒度。在用义素对概念进行定义时，所需义素的数目往往根据特定环境决定，所谓的“最小”也是相对的。如果是小范围的比较，义素所体现的颗粒度就粗些，所需的义素数目就可能少一些。如“男人—女人”可分解为：

男人——＋[男性]＋[成年人]

女人——－[男性]＋[成年人]

但如果加上“孩子”，则“成年人”须进一步分解，即：

男人——＋[男性]＋[成年]＋[人]

女人——－[男性]＋[成年]＋[人]

孩子——±[男性]－[成年]＋[人]

因此，在词义的构成中，各个义素所占的地位是不一样的。如上面的义素分析中，[人]和[成年]虽然同是义素，性质却不一样。[人]反映词语指示的对象的实质，±[男性]，±[成年]反映的是对象的性质和特征。所以有的学者在分析时对此作出区别，如贾彦德用“sh”这个符号表示“实质”，加在[人]这样的义素上，用“t”这个符号表示“特征”，加在±[男性]，±[成年]这样的义素上。如加上“sh”“t”等标志后，“女人”就应该分析为：

sh(human)t+(adult)t－(male)

西方有的学者把这种性质的符号称作标志（index）。“+”、“－”号也是这类符号。

在知网的建设中，义素的提取也综合运用了内省的方法和根据词典定义的方法。这大致可以分为两个步骤。第一步是对大约六千个汉字进行考察和分析以图提取出一个有限的义原集合。知网看到，中文中的字（包括单纯词）是有限的，并且它可以被用来表达各种各样的单纯的或复杂的概念，以及表达概念与概念之间，概念的属性与属性之间的关系。如下面的例子中，根据“治”“处”“理”三个字所表达的概念，得到了9个义原：

治：医治　管理　处罚　……

处：处在　处罚　处理　……

理：处理　整理　理睬　……

但其中有两对是重复应予合并。就这样，以事件类为例，知网在具有事件义原的汉字（单纯词）中提取出3200个义原，在初步合并后得到大约1700个，然后再进一步加以归类，便得到大约700多个义原。这是一个以汉字、词为根据的内省的过程。

第二步就是用归纳出的义原作为标注集去标注多音节的词，如果已有的义原不能区分有的多义词的诸义项所表达的概念时，标注集便需要进行合理调整或适当扩充。仍以事件类为例，经过标注，700多个义原就扩充为今天的800多个义原。在这个过程中，一些工具书对义原的提取起着不可或缺的作用。

知网列出了一些建设《知网》主要的，不可或缺的参考辞书，中国人民大学语言文字研究所编纂的《现代汉语通用字典》和中国社科院语言研究所词典编辑室编纂的《现代汉语词典》分列第一位、第二位。所以说，这也是个结合词典定义进行调整的过程。从这个意义上说，标注集的形成和知网的建设也是互动的。

另外，知网的主要目标是成为双语翻译的一个重要知识资源，所以语义颗粒度是比较细的。每个词条的DEF项（定义，意义的描写）的第一位置所标注的必须是知网所规定的主要特征，这一点很类似于贾彦德的"sh"所标识的"实质"。如"男人"定义中的"human|人"：

DEF=human|人，male|男

但是有些关系意义，可以把次要特征置于{}中后，作为第一位置标注。例如一些介词、连词等虚词，严格地说它们本身没有概念意义。

二、知网在应用义素分析上的优缺点

知网在应用义素分析中采用了一些灵活做法，避免了义素分析方法的一些固有缺点。

首先是定义中的常识和专业知识问题。义素分析可以帮助我们准确地掌握、解释、理解词义。但是，准确到什么程度，却是必须要考虑的。例如旧字典解释"鲸"为"海大鱼也"，但根据现代的生物分类知识，鲸并不是鱼，鲸表类别的义素应该是"+[哺乳动物]"。知网给出的定义是：

046543　鲸　beast|走兽，*swim|游

在对类别进行定位的同时，对特有的性质加以说明，即是"游"的主体。再如，对"男"的定义，有人分析成：

男:+[性别]+[体内产生精细胞]

女:+[性别]-[体内产生精细胞]

知网给出的定义则没有这种"学究气"：

061515　男　ADJ　aValue|属性值，sex|性别，male|男

其次是特征的非此即彼问题。最新发展起来的认知语义学是对义素分析的一个有力挑战。过去人们常认为一个事物或范畴的特征是有限的，非此即彼的，原始的，普遍的，抽象的，固有的（finite, binary, primitive, universal, abstract and innate）等等。而认知语言学的原型理论（prototype theory）认为范畴是原型的，成员是可扩展的，特征是文化的和特定语言的，是基于人的身体经验和后天习得的。

例如，"鸟"这一范畴的成员一般有以下基本特征：脊椎动物，卵生，有羽毛，有双腿和双翼，有喙，口内无齿，胸部有龙骨，会飞，会鸣叫……"鸟"的原型成员，如麻雀、鸽子、燕子、喜鹊、海鸥、乌鸦等，都具有上述几个特征。某些"鸟"的成员"会飞"的特征已退化，如鸵鸟、企鹅、鸭子等。但大多数鸟都会飞，而且，在日常语言的"通俗分类"（非科学分类）中，"会飞"应该算是鸟的一个相当典型的基本特征，也是鸟与其他动物相区别的重要特征。所以，"会飞"在一般人的语言意识当中应是

原型范畴“鸟”的一个特征。在知网中，对鸟的义素分析也照顾到了人们的心理印象，将鸟作为飞的主体：

bird|禽 [*fly|飞，~$consume|摄取，~?edible|食物]

为了避免前后矛盾，知网在具体的词语的定义时都加以说明，如：

066504　企鹅　bird|禽，^fly|飞

085670　鸵鸟　bird|禽，fly|飞

^表示不具有这个性质，或不能产生这种动作。这种定义显然不同于一般的义素分析的。一般的义素分析要求用绝对准确的量化成分构成来描写词义，哪怕只是极个别的鸟不会飞，那么“会飞”这一义素就不会包含在“鸟”的义素分析公式之中。这就是说，义素分析对某一词义义素构成的确定，采取“排中律”原则，或有，或无，可用严格的数学符号“+”和“－”表示，不能模棱两可。义素分析中不可能包含诸如“大部分成员都有的义素”或“成员基本上具有的义素”。而人的“通俗分类”中范畴可以包含某些非原型成员（如“鸟”中的“企鹅”）不具有的特征（如“会飞”）。所以知网是符合人们的认知规律的。

范畴的原型性给义素分析带来了极大的困难。如对于“母亲”，一般将其定义为：母亲＝生育者＋女性。但实际语言交际中“母亲”一词的用法与此并不完全相符。拉科夫认为，“母亲”这个概念至少牵涉到5个认知领域：1，遗传领域（the genetic domain），母亲提供遗传基因；2，分娩领域（the birth domain），母亲应生孩子；3，抚育领域（the nurturance domain），母亲应抚养孩子；4，谱系领域（the genealogical domain），母亲是最直接的女性祖先；5，婚姻领域（the marital domain），母亲是父亲的配偶。（Taylor1989:86）完全符合上述五个模式的是“母亲”的原型。除此之外，还有各种非原型的“母亲”：养母不符合1、2、4、5模式，继母不符合1、2、4模式，提供卵子给其他妇女的人不符合2、3、4、5模式，借用卵子生育孩子的人不符合1模式，等等。但这些妇女习惯上也被称为“母”，是“母亲”范畴中的非原型成员。

知网考虑到这些非原型成员的义素，引进了“家”的义素，归纳出了这些词语的共性：

060896　母亲　human|人，family|家，female|女

041696　继母　human|人，family|家，female|女

097421　养母　human|人，family|家，female|女

但由于采用了不严格的定义，所以和其它的词语不能区别开来，如：

062918　女儿　human|人，family|家，female|女

对“乳母”则采用了不对称的定义，也可以看出知网的灵活的特点：

071319　乳母　human|人，#occupation|职位，female|女，*feed|喂，employee|员

义素分析之所以能够流行，很大程度上是因为义素分析可以突出地显示词义之间的异同及联系。（王麦巧2002）例如常见的亲属词，军衔词等都可以按照义素的异同整齐地排成矩阵。其他如“狗、猪、鸡、鸭”……都可按此法分析。即使像知网无意于对词义进行区别的知识资源，根据它对主要特征的重视，

我们也能归纳出一个个相对完整的词语的集合来。

但是，由于知网并没有彻底地贯彻义素分析方法，所以，在规避了义素分析的一些固有缺点的同时，也丧失了这种分析方法的一些好处。

知网与一般的义素分析法最大的不同之处在于其并不注意词语之间的区别特征。例如，

067860　乔木　tree|树

033445　灌木　tree|树

005456　边疆　part|部件，%country|国家，edge|边

005466　边境　part|部件，%country|国家，edge|边

“乔木”和“灌木”的义素分析我们前文已经讨论过，但在知网中都只指示了其类别；“边境”，“边疆”都用来指靠近国境的地方，但是边境的范围小，边疆的范围大，如我们可以说“漠河是我国东北边境的要镇”，却不能说“漠河是我国东北边疆的要镇”。标准的义素分析法应该体现出这一点，但知网把它们混淆了。

再如，知网对上下义的标注也比较粗。如：

062463　鸟　　bird|禽

其下义词如“鹦鹉”“鸭”等的定义也都只指示了其类别。如：

118466　鹦鹉　bird|禽

一般的义素分析能够帮助我们说明词语组合的语义限制条件。语义限制条件首先表现为现实现象的关系，我们可以说“小王在看书”，“小姑娘在哭”，却不能说“书在看小王”，“愤怒的石头在哭”，因为现实世界的书、石头根本不可能发出看、玩、哭这样的动作，也不可能具备愤怒等感情，除非是在一个特定的体裁中，运用了特定的修辞手法。这些动作和状态都必须要求动作发出者的词语和中心语必须具备+[人]或+[动物]这样的义素。但根据知网，并不能实现这些限制。如“哭”，

051150　哭　weep|哭泣

查关于event的文件“概念的主要特征（1）”，weep|哭泣　{agent,target,cause|还是没有对动作的主体作出限制。

语义限制条件还表现为词语搭配对象的不同。例如“喝”要求的宾语一般须具有“[液体]”这样的义素，“粥”有“[液体]”这样的义素，所以“喝粥”能够成立。“饭”的义素是“＋[固体]”，“－[液体]”，所以“喝饭”不能成立。知网的定义是：

035714　喝　drink|喝

同样查“主要特征（1）”文件，知网也没有对动词的宾语作出限制。

如果义素分析很细，那么可以排除很多语义上的矛盾。如“张三的儿子怀孕了”是有矛盾的，因为“儿子”的义素中有＋[男]义素，而“怀孕”要求的主体有－[男]义素，所以就产生了矛盾。但这种矛盾根据知网是难以察觉的。知网对“怀孕”的定义是：

037951　怀孕　pregnant|怀孕，#medical|医

“怀孕”在“主要特征（1）”文件中，也仍是作为一个义原处理的，没有对其主体的语义类型作出任何限制。

从以上的分析可以看出，知网并非是像Cyc那样的一个包罗万象的知识库，它虽然采用了义素分析的方法，但是并不严格，缺乏很多必要的区别性信息。这是由它的建设目的决定的。知网主要是为了机器翻译的目的建设的，对于英语和汉语之间的概念的联系作了详尽的考虑，由于采用了义素分析的方法，所以也能在一定程度上揭示一个语言内部的概念之间和概念的属性之间的联系。但是，如上所述，它兼顾到英汉两种语言，很多概念对于另一种语言是冗余的，所以人为地制造了很多歧义。如“草包”一词，在知网里被列为8个词条，其实在汉语里可归并为两条：

009899　草包　　N　human|人，unable|庸，undesired|莠　blockhead

009905　草包　　N　tool|用具，cubic|体，@put|放置　　straw bag

知网为good-for-nothing, idiot, imbecile, nincompoop, nitwit另立词条，定义与009899条相同；为straw sack另立词条，定义与009905相同。这样，知网就等于把两个语言的歧义叠加在一起，这给后续处理的歧义消解带来极大的困难。再加上它没有系统地采用义素分析方法，对概念之间的选择限制没有进行深入的探讨，这在一定程度上限制了它在语义分析领域的应用。

三、语义场理论和义素分析的相关性

特里尔（Trier）在1930年提出的语义场理论被广泛地认为是“开创了语义学史上的新阶段”（Lyons1977：250）。在很长时期内，人们对多种语义场（如对颜色、亲属、官衔等语义场）作了详尽的描写。在描写过程中，人们也发现对所描写的语义场作出准确的界定是十分重要的。但界定的标准牵涉到两个方面。一方面要确定一个语义场在整个语言的更为宏观的语义场的层次结构中的位置，另一方面也要确定所要研究的语义场自身的规模，只有这样，才能对语义场进行比较科学的描写。

例如要研究汉语中的“亲属语义场”，首先就需要对这个语义场进行定位。亲属语义场是“人”这个语义场下的子场。因为“人”语义场还下分很多语义子场，亲属语义场只是从“亲属”这个角度对“人”进行语义场的归类。在确定好语义场的层次位置后，还要在微观上对语义场进行考察。语义场的基本构成单位应是概念，词语也应围绕着概念进行聚合；现代汉语的亲属词有几百个，但许多亲属词几乎是同义的，在进行语义场分析时，首先要对这些词语进行聚类。例如，与“父亲”同义的亲属词就有“爸爸”“翁”“爹”“老子”“阿爸”“家父”等，与“岳父”同义的亲属词有“岳丈”“岳翁”“丈人”“外父”“泰山”“冰翁”等，我们就以“父亲”，“岳父”作这两个同义集合的代表。这样经过筛选，可以得到亲属语义场大致有五十三个同义集合构成，对这五十三个概念或同义集合进行进一步的聚类和层次分析，“亲属语义场”的描写就比较充分了（郭伏良1995）。在对语义场进行定位的过程中往往牵涉到范畴的界限问题。例如，要研究英语中dwelling place这一义场，就首先要搞清楚这一义场是否只是指人的dwelling place, 如:house，apartment，villa，chateau，hut,cabin等，是否包括动物的dwelling place，如:nest,cage，stable，pigsty等。如果是人的dwelling place, 是否包括一些临时性的dwelling place, 如

:hotel, inn, shelter, asylum等和一些或多或少与dwelling place有关的地方，如:hospital, prison等，由于范畴的原型性，这类问题有时比较棘手，需要根据具体的应用确定描写的范围。

为了能准确界定一个语义场，人们深入分析词汇的语义关系，甚至进一步深入到了词义内部结构，这直接导致了义素分析方法的提出，也反过来促使了人们在一个更加具体，清晰的层面对语义场进行定义：语义场实质上是由具有某些共同义素的词群聚合而成的场。人们进一步对语义场的层次性进行重新观察，发现某些词可以在一个共同概念或语义成分的支配下形成语义场。表达共同概念的词称为上义词（superordinate），语义场就由上义词和若干个受上义词支配的下义词（hyponym）组成。如creature一词，作为上义词，与其支配的bird, fish, insect等一系列与其具有共同语义特征的下义词一起组成“动物场”；再如plant一词，与受其支配的flower, tree, crops等词构成“植物场”。作为上义词的creature和plant又可受living thing的支配，转而变成两个并列下义词，一起构成更高一层的“生物场”。上义词和下义词具有明显的相对性。如以上bird, fish和insect三个词，一方面，相对于creature而言是下义词，另一方面相对于lark和eagle, cod和trout, butterfly和ant等词而言又都可以成为上义词，构成各自独立的语义场。属于同一个语义场的词在语义上相互依存，相互制约。也就是说要确定某个词的意义，必须首先比较该词与同一语义场中其它词在语义上的联系，以及该词在语义场中所在的位置。

四、WordNet中的语义场理论

WordNet就是主要根据上述语义场理论实现的一部机器词典。虽然WordNet并不明确承认，但根据其建设方法和词典的形态可以看出这一点。建立包含词语意义描述的大规模词库的常规方式是语义成分分析（即义素分析法）。WordNet主帅George.A.Miller在1976年与Philip N. Johnson-Laird合作的*Language and Perception*一书中也踌躇满志地探索义素分析的语义描写方法，但是要定义一套原子概念却并非易事，直到1985年，他们仍然没有能够拿出一个完整的定义清晰的原子概念的清单。

到1985年，许多认知心理学家和计算语言学家开始以“网”的形式来描述词语的意义。比如：“桌子”（table）和“家具”（furniture）代表两个节点（node），而这两个节点之间有一个箭头（dart）来表示这样的命题：桌子是一种家具（a table is a kind of furniture），即“Is-A-KIND-OF”这样的语义关系。随着这方面研究的增多，越来越多的人自觉地意识到：除了利用语义成分（义素分析法）表示语义，还可以利用关系来表示语义（基于关系的词汇语义学relational lexical semantics），而且后者有可能替代前者。基于这个理念，从1985年开始，WordNet作为一个知识工程全面展开。

WordNet以同义词集合（synset）作为基本建构单位（building block）进行组织的，这也正体现了构成了语义场理论的基础的信念，即一个给定语义场中的词的意义来自于它跟该语义场中其他词之间的相似或相反关系。在WordNet中，用户脑子里如果有一个已知的概念，就可以在同义词集合中找到一个适合的词去表达这个概念。WordNet中的基础语义关系是synonymy（同义关系）。

在词汇语义学的理论中，如何用定义来表示词语化的概念，取决于该理论是打算成为构造性的还是仅仅是区分性的。知网变通运用义素分析法这种方法，体现了构造性的特点，但不能有效区分很多词

的意义差别，区分性不强；WordNet恰恰是用区分性理论来表示词义的。如果阅读该定义的人已经获得了该概念，仅仅需要辨认它，那用一个同义词（或近义词）通常就足够了。例如，某人已经知道board可以指称一片木材或是一群为了某个目的集合起来的人，只需要得到plank或committee的帮助就可以挑出原义。同义词集{board,plank（板材）}和{board,committee（委员会）}可以作为board这两个义项的无歧义的指示器。由于英语中同义词很多，synsets通常足以用来作区分的目的。所以说对区分性的理论的要求比较低，但一般也足以构造出所需的映射。但有时候，也可能找不到一个合适的同义词，例如，board的另一个义项是“包伙食”，这时WordNet用一个很短的注释来解决这个多义问题，如{board,（包伙食，即定时提供一个人的三餐以赚钱）}可以用来区分board的这一意思，它可以被看成一个只有一个成员的synsets。

WordNet将词汇分成五个分类：名词，动词，形容词，副词和虚词，实际上，WordNet仅包含名词，动词，形容词和副词。WordNet包含了大致57,000个名词词形，它们被组织成大致48,800个词义（synsets，同义词集合）。另外还包含11500个动词同义词集合和16428个形容词同义词集合，包括许多名词、分词（participles）和介词词组。这些数字都只是一个约数，因为WordNet还在不断地扩充。同义集合看上去数目非常庞大，且比较零碎。但是，这些同义词集合都是有组织的。名词在词典存储中是按主题的等级层次组织的，动词按各种搭配关系来组织，形容词和副词以N维超空间组织。这就决定了WordNet最终体现了语义场的理论。

WordNet名词是按照层次体系组织的。组织的方法是从分成一组组的语义成分（类似于义素）中选出一定（相对较小）数目的一般化概念，作为每一个单独的层次体系结构的唯一起点。这些层次体系结构均对应于相对独立的语义场。语义场中所有单词的原始语义成分均来自那个表示一般化概念的唯一起点。以下就是WordNet名词的25个语义场的唯一起点：

{行为，动作，活动}{食物}{所有物}{动物，动物群体}{组，分组}{过程}{人工制品}{位置，地方}{数量，量}{属性}{动机}{关系}{躯体特征，身体}{自然物体}{形状}{认知，知识}{自然现象}{状态，情况}{通讯}{人，人类}{物质}{事件}{植物，植物群体}{时间}{感觉，感情}

类似地，WordNet中的动词也分成了身体动作动词（Verbs of Bodily Functions and Care，275个同义词集合），变化动词（Verbs of Change，约750个同义词集合），通信动词（Verbs of Communication，710个以上的同义词集合），竞争动词（Competition Verbs，200个以上的同义词集合），消费动词（Consumption Verbs，130个同义词集合），接触动词（Contact Verbs，820个同义词集合），认知心理动词（Cognition Verbs），创造动词（Creation Verbs，250个同义词集合），运动动词（Motion Verbs，500个同义词集合），情感心理动词（Emotion or Psych Verbs），状态动词（Stative Verbs，约200个同义词集合），感觉动词（Perception Verbs，约200个同义词集合），领属动词（Verbs of Possession，约300个同义词集合），社会互动（Verbs of Social Interaction，约400个同义词集合），气象动词（Weather Verbs，约66个同义词集合）等15个基本语义类。

但WordNet不仅仅是用同义词集合的方式罗列概念。同义词集合之间是以一定数量的关系类型相

关联的。这些关系包括上下位关系、整体部分关系等，不同词类中的语义关系类型也不同，比如尽管名词和动词都是分层级组织词语之间的语义关系，但在名词中，上下位关系是hyponymy关系，而动词中是troponymy关系；动词中的entailment（继承）关系有些类似名词中的meronymy（整体部分）关系。名词的meronymy关系下面还分出三种类型的子关系，即部分、成员和材料等。反义关系为WordNet中的形容词和副词提供了一个中心组织原则。

我们以兼名，动，形，副四类词的多义词long为例来说明WordNet的体例。名词有一个义项：

1. {10951384} <noun.time> long#1 -- (a comparatively long time; "this won't take long"; "they haven't been gone long")

动词有两个义项：

1. {01247460} <verb.emotion> hanker#1, long2#1, yearn#1 -- (desire strongly or persistently)
2. {01231785} <verb.emotion> long1#2, ache6#2, yearn1#2, yen#1, pine#1, languish#2 -- (have a yen for)

形容词有10个义项，由于篇幅的原因，我们只取前两个义项：

1. {01380813} <adj.all> long2#1 -- (primarily temporal sense; being or indicating a relatively great or greater than average duration or passage of time or a duration as specified; ; "a long life"; "a long boring speech"; "a long time"; "a long friendship"; "a long game"; "long ago"; "an hour long")
2. {01376707} <adj.all> long1#2 -- (primarily spatial sense; of relatively great or greater than average spatial extension or extension as specified; "a long road"; "a long distance"; "contained many long words"; "ten miles long")

副词有2个义项：

1. {00162139} <adv.all> long#1 -- (for an extended time or at a distant time; "a promotion long overdue"; "something long hoped for"; "his name has long been forgotten"; "talked all night long"; "how long will you be gone?"; "arrived long before he was expected"; "it is long after your bedtime")
2. {00162432} <adv.all> long3#2 -- (for an extended distance)

我们看到，每个义项基本上都有注释。这主要是随着WordNet中词语数量的增加，要想既尽可能清楚地区别词义，同时又要保持同义词集合的纯粹性，仅靠同义来定义词义就越来越困难了，所以还要以注释加以补充。一开始，WordNet保持注释尽量得短。但后来，注释就变得越来越长，数量也稳步增长。

这种方法正说明了WordNet的区分性而非构造性的特点。它是在宏观的角度把握词义，把词义当做一个整体，在语义场中考察词语。这类似于中国的传统的训诂学和传统语义学的解释词义的方法，用一个词去直训另一个词，把词义当作一个囫囵的整体。如《尔雅・释诂》把“崇”解释作“充也”，《说

文·言部》把“讯”解释作“问也”等。就是训诂学家用一个词组、句子，甚至用一个以上的句子解释词义，也是把被解释的词义当作一个整体，而不把词义看作是由若干成分组成的。例如《尔雅·释天》“谷（穀）不熟为饑，蔬不熟为馑，果不熟为荒。”“饑”这个词的词义是什么？《尔雅》指出就是“谷不熟”这样一个词组的整体的含义。又如《说文·山部》：“岑，山小而高。”这就是说“山小而高”这个句子的含义就是“岑”的词义（贾彦德1999：51）。

五、结语：回顾与展望

在语言学和计算语言学中，“分解”的思想产生了迄今为止影响最大的语义处理方法，即义素分析方法。语义分解的理论和方法之实质是利用某些任意的“义素”或“语义特征”来描述意义的深层结构。从理论上说，如果有足够的“义素”，就可以描述所有词的全部意义。然而实际上要确定一个词究竟含有多少成分，含有哪些成分，是很困难的。这是因为相对于形式的句法分析，语言的“意义”总是显得更为模糊，难以确定。除此之外，各人对同一个词的理解不同，很难制订出统一的义素或语义特征。同样是用义素分析的方法，对“父亲”的义素分析就有以下几种：

父亲:+[有子女]+[成年]+[男人]（徐思益1994）

父亲:+[男性]+[直系亲属]+[长辈]（王德春1983）

父亲:（亲属）→（生育关系）+（男性）（贾彦德1999）

DEF=human|人, family|家, male|男（知网）

因此很多人据此认为，意义的本质是不可分割的，对一种不能分割的东西采用分割的办法来处理，结果可想而知；所以，义素分析技术本身是难有出路的，至少是不能完全处理语义问题的。但是，当人们看到一个方法的局限性并对它进行不遗余力的批评的时候，却常常把问题也丢了。比如义素分析有很大的局限，那么语义特别是词汇语义如何被描写，仍然是个问题，问题没有实质性解决。义素分析方法是有历史贡献的，它告诉我们，语义是可以被把握，被描写的，是可被计算的。在没有更加成熟的方法之前，义素分析方法还是不失为一个好的方法。

而关于语义场理论，其实有强弱之分。特里尔的理论是语义场理论的“强”版本。特里尔着眼于概念空间，并从概念出发，从概念的角度研究词。他认为，概念场是围绕中心概念组成的相互联系的概念的广泛系统，在语言中有词场与之相对应。他认为词汇场是一种覆盖物，就像马赛克一样覆盖于概念场之上并将之切分。一个概念场可以被几个词汇场所覆盖（寇永良2001）。这种版本实际上把一个语言的词汇V看作是词（lexemes）封闭的集合，$V=\{l_1, l_2, l_3, \cdots\cdots l_n\}$，并可以进一步划分为一个（词汇）语义场（Lexical Fields）的集合$\{LF_1, LF_2, LF_3, \cdots\cdots LF_m\}$。其中，任意两个场的交集都是空集；所有的场的并集就等于V（Lyons1977：268）。但是，词汇和词汇中的每个语义场都是词的封闭集合显然是不妥当的。大部分语义场都不像特里尔所说的那样界限分明，结构严密，仅仅是二维的平面的东西。词汇和语义场都至少是一个开放的集合，应该定义为$V=\{l_1, l_2, l_3, \cdots\cdots l_n \text{ etc.}\}$，$LF_i=\{l_{i1}, l_{i2}, l_{i3}, \cdots\}$。

典型的语义场分析主要考虑聚合关系（paradigmatic），有的分析也考虑组合关系（syntagmatic），

如波尔齐格（Porzig）举例说英语中的blond和hair，dog和bark，solve和problem总是固定搭配一起，可以认为这种词义的组合关系也构成语义场。这种组合的语义场常被看作是对聚合的语义场的一个重要补充。（Lyons 1977：261）在WordNet中，关于一个词的选择优先性的信息不是网络结构的一部分，这种信息往往出现在跟一个同义词集合（synset）相伴随的注释中，例如，动词同义词集的注释信息常常具体地标示什么类型的名词是作为该同义词集中动词的典型论元出现的。实际上，动词的十几个类别就是根据搭配的论元的类型划分出来的。

在汉语信息处理中，像知网这样的基于构造性的知识资源已经建成了，但仍缺乏WordNet这样基于区分性的知识资源。由于这种知识资源独特的应用前景，已经有数个单位如东北大学在着手将WordNet进行本地化。（靳光瑾2001：40）从概念角度来说，语言之间有相通性，上下位语义和树形关系、继承关系都有可移植性。但是考虑到语言空间，英语和汉语还是有巨大差异的，同样是词，汉语中的词就没有明确的形式标志，无法和英语的词进行一一对应。因此，本地化可能不是一般的修改或扩充就可以了，可能是要从根本上重建（靳光瑾2001：44）。这无疑是一项浩大的工程。如果考虑同时在工程中综合应用包括语义场理论在内的多种现代语义理论以对WordNet作出改进，那无疑更增加了工程的难度和挑战性。

参考文献

陈群秀 1998《一个在线义类词库：词网 WordNet》，《语言文字应用》第 2 期。

董振东，董强1999《知网简介》，www.keenage.com。

董振东，董强2001《知网和汉语研究》，《当代语言学》，第 1 期。

郭伏良1995《现代汉语语义场分析初探》河北大学学报（哲社版）第 1 期

Allen, James1995 *Natural Language Understanding*, The Benjamin/Cummings Publishing Company, Inc.

贾彦德1999《汉语语义学》，北京大学出版社。

靳光瑾2001《现代汉语动词语义计算理论》，北京大学出版社。

寇永良2001《语义场浅释》，《黑龙江教育学院学报》第 2 期。

Lyons, J. 1977 Semantics , Cambridge :Cambridge UP.

孟琮等1987《动词用法词典》，上海辞书出版社。

George A. Miller, R.Beckwith, C.Fellbaum, D. Gross, and K. Miller 1993 Introduction to WordNet: An On-line Lexical Database, in *Five Papers on WordNet*, CSL report,Cognitive Science Laboratory, Princeton University.

Taylor,J.R. 1989 *Linguistic Categorization: Prototypes in Linguistic Theory*, Clarendon Press: Oxford

王德春1983《词汇学研究》，山东教育出版社。

王麦巧2002《义素分析刍议》，《渭南师范学院学报》第4 期。

吴立德1997《大规模中文文本处理》，复旦大学出版社。

伍谦光1988《语义学导论》，湖南教育出版社。

徐思益1994《论句子的语义结构》，《语义学论文选》，新疆大学出版社。